Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

A Cloud-Based Deep Learning Framework for Real-Time Anomaly Detection in Public Surveillance Videos Using VGG16-LSTM

Sonia Chourasiya*[1], Dr. Pharindra Kumar Sharma [2]

[1*]M.Tech Student, Dept. of CSE, SRCEM
[2]Associate Professor, Dept. of CSE, SRCEM
*Corresponding Author.: chourasiyasonia24@gmail.com

Abstract

The purpose of this research is to design and train a deep learning-based surveillance network that can detect anomalous and suspicious behavior in real time using the Daily Crime Surveillance and Safety System (DCSASS). The system will enhance safety on the streets and enable detection of threats ahead through automation of anomaly detection of the video feeds. The methodology is systematic and organized and starts by gathering and authenticating video data of a high risk activity of 13 high risk classes of activities such as assault, robbery, arson and vandalism. An effective preprocessing pipeline was developed, including a Gaussian blur filter and brightness-contrast normalization to improve clarity of the frame and reduce noise. The videos were then split into frames and annotated by using official labels and separated into training and validation sets using stratified sampling to ensure class balance. A hybrid architecture was followed whereby the VGG16 convolutional neural network that was used as the frozen spatial feature extractor was integrated with the Long Short-Term Memory (LSTM) layer to learn temporal interrelations. The time intervals of the input feature vectors were converted to a repetition, simulating the motion context. The performance, such as the accuracy, loss, precision and recall, were modelled using the standard metrics. The proposed VGG16-LSTM model presented an accuracy of 95.05, low loss of 0.126, precision of 95.46 and recall of 93.11 indicating that this model is very reliable when it comes to detecting anomalies. In addition, object detection analysis yielded a precision of 85.7% and mAP 0.5 of 71.7%. The validity and scalability of the suggested framework were validated in comparison to the other existing models, including ResNet-50 and an earlier hybrid model.

Keywords: Cloud Computing, Public Security, Video Surveillance, Image Analysis, Anomaly Detection.

Introduction

There has never been a greater need to have good surveillance systems as cities continue to expand rapidly and people relocate into them. The old methods of video surveillance are useful, but they frequent fail in storage space, processing capabilities, and the capability to process data as it occurs. Conventional answers are becoming less and less useful as urban areas continue to become more complex and as the number of public security cameras gathers more and more data at an alarming pace. Cloud computing is a technology which is game-changing and which addresses these challenges by providing us with scalable infrastructure,

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

advanced data processing and real-time analytics. The cooperation of cloud computing and public security video image investigation systems is a giant leap towards police. It facilitates ease of crime discovery, quick response to situations and ensure safety of the people by being proactive. A cloud-based image investigation system of a public safety video surveillance integrates the best of both smart video surveillance and the speed and flexibility of cloud computing. These systems are able to record, archive and scan video of thousands of surveillance cameras installed in towns and cities. This provides the police with one place where they can handle all their security operations. Cloud-based resources can provide investigators with high-definition video footage wherever they go. They are also capable of locating objects, identifying individuals, tracking movements and locating unusual objects through sophisticated algorithms without needing servers or hardware in the vicinity[1]-[3]. This reduces the costs of operation as well as it simplifies the large scale monitoring activities as they become more accessible, dependable and scalable. Cloud computing also dynamical enough to allow the change of the public safety systems with changing technology. The introduction of artificial intelligence (AI) and deep learning algorithm to cloud-based systems, e.g. has enabled the automatic evaluation of movies, as never before, with greater accuracy. These are some of the skills that would be very helpful in observing suspicious behavior, recognizing people wanted and reporting directly to the police. Cloud-computing systems are capable of rapidly reviewing live video feeds, marking important occurrences and coordinating rapid responses to events such as riots, accidents or terrorist attacks. This not only assist the police and other law enforcing agencies to be aware of what is happening, but also makes the citizens feel safer knowing that the security measures of the government are at work. The cloud-enabled security systems are also centralised, which facilitates agencies to collaborate and exchange information. By storing video data in the cloud, various police departments, forensic teams, and government organisations can collaborate on investigations by providing them access to synchronised and time-stamped[4]–[6] footage. This eliminates data silos and assists in creating crime databases that communicate with one another. They are required to monitor crime in various locations and come up with models that can help to stop crime. Strong encryption and access controls are also part and parcel of cloud computing to ensure that private surveillance data cannot be accessed by hackers and other individuals who do not have the right to access the data.

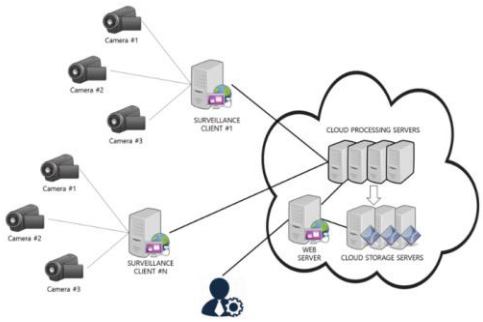

Fig. 1 public security video image investigation [7]

The other significant benefit of cloud-based public security systems is that it can be used to assist in real-time data analysis. Intelligent surveillance cameras can also handle video data locally and transmit it to the cloud via such such technologies as edge computing and the Internet of Things (IoT). This decreases latency and allows the cameras to be more rapid. This mixed-design will only transmit valuable or filtered information to the cloud to be analyzed in detail. This saves bandwidth, and speeds up decision-making. As an example, traffic cameras with connectivity to cloud-based software can assist in managing traffic in a city and ensuring road safety by identifying traffic violations, tracking the movement of cars on the road, and predicting when traffic will become heavy. Moreover, cloud computing simplifies the establishment and expansion of surveillance systems [8]–[11]. Cloud-based systems allow the growth of businesses by public safety agencies as they need them since they need not purchase real servers and hire specialised maintenance

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

personnel which is an expensive affair. Rather, they buy what they require. With modest service elements and low-cost cloud subscriptions, even small towns, rural regions, and poorly developed nations with no greatest surveillance frameworks can obtain secure and useful observation. Security resources can be distributed more equally across various locations by providing monitoring technologies to all. Integrating cloud computing in the video systems of the police, however, has a number of problems. You must be aware of how to discuss matters that arise concerning data privacy, within the law, and ethical surveillance. Individuals must have the capability to have confidence in the government and there should be definite legislations on the way information is collected, stored and utilized to ensure that it is not misused or abused. Governments and businesses that access a video surveillance as a source of personally identifiable information (PII) should also observe the local and international regulations such as GDPR. Of special importance is the need to protect civil rights in this scenario and make the best out of technology by employing anonymization methods, audit trails, and regulatory controls [12], [13]. The establishment of a public security video image investigations system by the use of cloud computing is a monumental shift in the way societies conceptualize crime prevention, surveillance and security. It takes advantage of the fact that the cloud is able to circumvent the issues of the older systems and enables real-time video analysis, scalable data management, multi-agency collaboration, and intelligent decision-making. These systems are significantly improved when they operate with AI, IoT, and edge computing [14], [15]. They can adapt to new security needs and be more powerful. There is much potential in making communities smarter, safer and more resilient in the digital age through cloud computing, but there remain issues that must be addressed, particularly in regards to data governance and accountability to the people[14], [15].

2. **Literature Review**

Hamdaoui 2020 et al. IoT is altering the way we think, act, and live by making services that weren't possible before. In smart cities, it helps with things that need to happen right now, like watching videos, controlling traffic, and responding to emergencies. These rely on a number of IoT devices that deliver data and work together to process it so that all the facts are available and decisions can be made swiftly. Semantic Virtual Space (SVS) is a concept that was created to meet these needs. It is an abstraction made to be executed in the cloud and virtual machines in the IoT infrastructure. Scalable structures and processes are suggested to be deployed and managed to scale the deployment and management of numerous SVS instances. This would help smart city applications to serve their evolving needs in a fast and reliable manner [16].

Toussaint 2020 et al. Machine learning systems are becoming more and more significant in the Internet of Things since they aid with edge intelligence and the trend towards intelligence everywhere. Both areas have made progress, but it's hard to combine them because the IoT is decentralised, has a lot of devices, and doesn't have a lot of resources. Traditional ML systems are designed for huge cloud environments, which are significantly different from what IoT needs. To address this, recent research has been devoted to the distribution of ML on the cloud, edge, and IoT levels. People view edge intelligence as a socio-technical system, and therefore should be constructed considering the needs of all users, system constraints and trade-offs. The researchers should strive to develop edge intelligence systems that are reliable and can be expanded in the future [17].

Shao 2020 et al. it is very important to be able to keep your eye on ships, when they are in the vicinity of a coast, and to control the maritime traffic and ports. Previously, remote sensing or radar photos have not been of much use. T is able to find things in real time using pictures that are captured by security cameras on the ground. Yet it is difficult to be precise and speedy when there are rich backgrounds and numerous kinds of ships. We offer a new cognisant saliency-aware CNN architecture. It makes things work with the help of coastal priors, deep features, and saliency maps. The model involves using CNN to predict the types and locations of ships, and salient region detection is used to repair the locations. On Darknet, using CUDA, the model is more efficient and quicker than Faster R-CNN, SSD, and YOLOv2 [18].

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

Kitchin 2020 et al. Contact tracing, quarantine enforcement, and symptom tracking were instantly implemented with the help of surveillance technology such as smartphone apps, facial recognition, infrared cameras, biometric wearables, drones, and predictive analytics against COVID-19. The priority of public health over civil liberties frequently necessitated speedy action. We are yet to find out the value and usefulness of these technologies. There is concern among people on how they will impact privacy, government control and the emergence of surveillance capitalism. Such techniques may not just be effective in preventing the virus, they may also increase the invasive surveillance. This may influence the way the government and public health operate in a manner that is dreadful to all [19].

Shah 2020 et al. Smart devices like phones, watches, sensors, and on-board units make life easier, but also use a lot of energy, which affects the environment, the cost, and the lifespan of the equipment. A multi-level, fog-based framework that saves energy is suggested for smart settings that use the Internet of Things (IoT). It adds two layers: a policy layer and energy-efficient technology. These layers keep track of usage and help people make decisions that are good for the environment. Framework finds energy sources, evaluates the needs of each task, looks for low-energy options, and picks devices that do the job well. The proposed approach has been tested in smart parking, ICU management, agriculture, and airports, and simulations with iFogsim show that it can save a lot of energy[20].

Table: 1 Literature Summary

Authors/year	methodology	Research gap	Findings
Peng/2020 [21]	Digital Twin-based hospital management methodology.	Lack of real-time integration in hospital facility management systems.	Improved efficiency, reduced energy use, fewer faults, and enhanced maintenance quality.
Yadav/2020 [22]	Automated COVID safety monitoring methodology.	Lack of real-time automated monitoring for mask and distancing compliance.	Real-time system effectively detected mask and distancing violations, reduced manpower.
Henman/2020 [23]	AI-driven public administration enhancement methodology.	Limited understanding of AI's challenges in public sector governance implementation.	AI improves public services but raises issues of bias, accountability, legality.
Gupta/2020 [24]	IoT-based pandemic response framework methodology.	Lack of integrated IoT frameworks for pandemic prevention and management.	IoT-enabled systems enhance pandemic monitoring, prevention, control, and community resilience.
Ali/2020 [25]	Component-based cloud testing framework methodology.	Inefficient test case prioritization in dynamic cloud-based component systems.	Proposed framework improved fault detection rate over existing prioritization methods.

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

3. Research Methodology

Employing the DCSASS dataset, this study follows an organised process to create a surveillance system based on deep learning. It starts with gathering and checking data to make sure that the class is represented correctly. Blurring and normalisation are two preprocessing methods that improve video quality. After that, frame extraction and annotation mapping are done using official label files. Then, using stratified sampling to keep the class balance, the dataset is partitioned into training and validation sets. Before being sent to a VGG16-LSTM hybrid model, each frame is shrunk and made normal. This model finds problems by using both spatial and simulated temporal characteristics. The method makes sure that the data is accurate, can grow, and stays consistent.

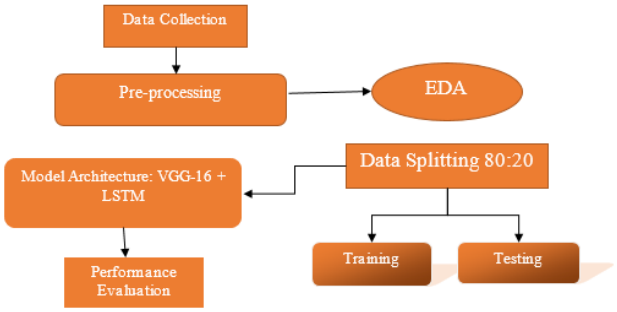

Fig. 1 Proposed flow chart

A. Data Collection: Daily Crime Surveillance and Safety System (DCSASS) is a comprehensive, and diversified source of real-world video surveillance information which was specifically created to be used in smart security applications. It also consists of 13 very important and destructive action groups which are as follows: abuse, arrest, arson, assault, burglary, explosion, fighting, theft, road accidents, robbery, shooting, stealing, and vandalism. These groups are based on the real life dangers and oddities in circumstances of communal security. In order to ensure that the dataset was useful and in the right format, a bespoke Python script was created to analyze the raw video files and sort them. This script removed redundant folders (such as label folders), left only valid top-level action categories, and then repeated the process of subfolders searching and counting video files by format. It also verified the existence of crucial data structures and reported any discrepancies, lost files or naming mistakes. By having clear outputs to the console, the script allowed one to easily determine which files were being used and the number of videos in a particular class. This rigorous validation procedure ensured that the resulting data set was clean, consistent and well-organized. This allowed it to be easily incorporated into automated preparation steps and applied to machine learning methods.

B. Data Preprocessing: To preprocess the videos to fit in downstream modeling and achieve optimal results in classification tasks, a detailed and efficient preprocessing pipeline was created. The purpose of this pipeline was to make each frame visually clearer, make the resolution of each frame the same throughout the dataset, and minimize background and environmental noise which would otherwise interrupt feature extraction. Each video was frame by frame processed with a series of improvement algorithms. First, Gaussian blurring was used to eliminate high-frequency noise and interferences at pixel level. Contrast and brightness normalization followed, which was done through the use of OpenCV cv2.convertScaleAbs are used with alpha=1.2 and beta=15 which served to normalize variations in illumination and enhance the conspicuity of action-relevant features. The frames were then dynamically coded as high-quality mp4 files with the help of cv2 after

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

improvement. Video Writer, which guarantees compatibility with subsequent processing stages. The whole pipeline took advantage of the multiprocessing module in Python to run simultaneously, allocating video processing jobs to all the available CPU cores. This method greatly improved the preprocessing process, and it could be applied to large data sets. The uniform visualization and increased clarity obtained in this way significantly improved the downstream model in identifying complex patterns and exposing anomalies in surveillance video, thereby providing a solid base of dependable action categorization.

C. EDA: Before training the model, we conducted Exploratory Data Analysis (EDA) to get to know the structure, distribution, and quality of the DCSASS dataset. The dataset has 13 action classes, and EDA revealed that certain action classes were more prevalent in the dataset than others. Binary labels (normal vs. anomalous) distributions were also considered and it was found that anomalous cases were a bit more in number than those that were normal. We verified the consistency of the images and labels on frame-level and identified any missing or mismatched data entries. Metadata of the video such as its resolution, length, and frame rate were examined. EDA was used to ensure the data was accurate, aided in judgements concerning preprocessing, and provided a solid foundation to create the accurate models.

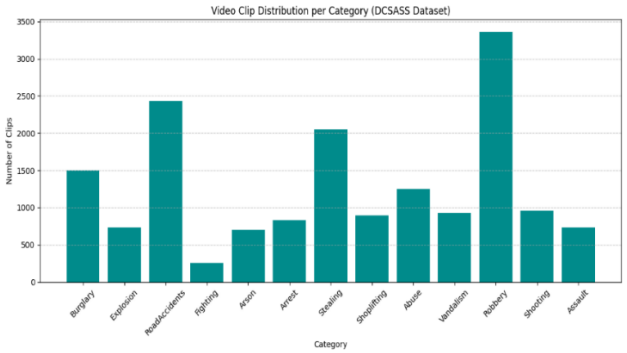

Fig. 2 Video clip distribution by category

Fig.3 shows the distribution of video clips under the 13 categories of action in the DCSASS dataset. The imbalance in classes is also emphasized in the visualization, as some categories (Fighting and Shoplifting) have much more clips than others (Explosion or Shooting), showing the importance of having stratified sampling during model training.

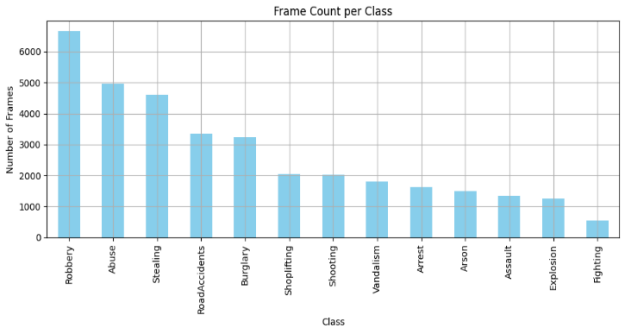

Fig. 4 Frame count per video class

Fig.4 shows the number of frames per video class in DCSASS dataset where there is a difference in the number of extracted frames per category. Classes such as Assault and Robbery are represented with more frames whereas classes such as Explosion are represented with less frames which depicts irregularities in classes, hence could affect the model learning.

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

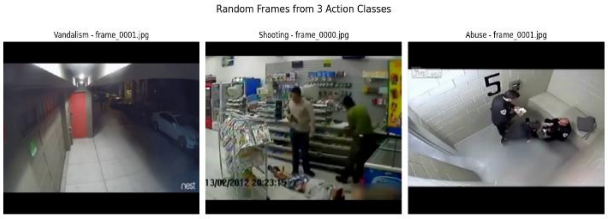

Fig. 5 Random frames from action classes

Fig. 6 shows random samples of frames in different action classes in the DCSASS data set. These samples indicate that there is a variety of visual scenes, lighting conditions, and types of activities in categories. The pictures illustrate distinct characteristics of the classes, which confirms that the dataset can be used to train deep learning models in action recognition in the context of surveillance.

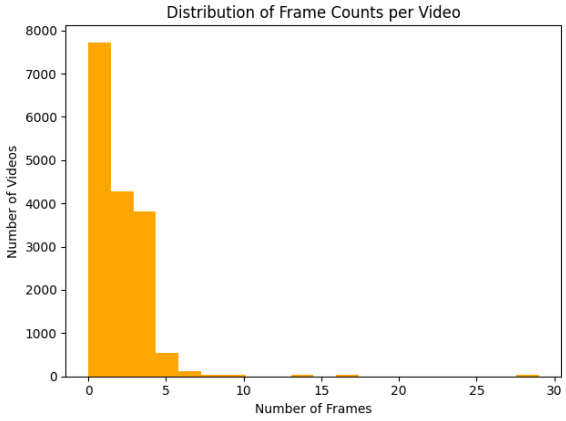

Fig. 6 Frame count distribution per video

Fig.6 depicts the number of frames per video in the DCSASS dataset, which indicates the number of frames that were identified in each clip. The distribution shows that some videos have considerably more frames than others, thus necessitating uniform sampling to balance the datasets during training.

D. Data Splitting: Training the YOLO object detector model required a well-organized dataset of annotated frames to be divided into training and validation parts and adjusted to the format that the YOLOv5 requires the dataset to have. This started by stratified sampling to make sure that in both subsets every action class was represented proportionally. In case the dataset was more than 10,000 frames, it was down sampled randomly and the class balance was maintained so that the training efficiency was optimal. The train_test_split was done with stratification on the action class labels to divide the data into an 80:20 split that guarantees a balanced representation of categories of objects in the training and validation sets. All chosen image frames (.jpg) and their annotation files (YOLO .txt) were transferred to images and labels folders of corresponding train/ and val/ directories. Such a directory structure is required to be compatible with YOLOv5 and other YOLO-based frameworks. The data splitting script had strong validation in place, which ensured that both image and label files are present and copied. In case of the absence of a corresponding label or image, the script recorded informative messages to assist in the error tracking and data integrity. Such attention to detail and error correction allowed making the dataset clean, structured, and prepared to train on object detection.

E. **Feature Extraction:** All video frames extracted were initially processed and resized to standard size of 128x128 pixels with three color channels of RGB to achieve standard input dimensions throughout the dataset. This resizing played an important role in compatibility with deep learning architectures as well as to standardize the visual size of features. The pixel values were then scaled to

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

[0, 1] by dividing them by 255, a widely used trick that helps increases the stability of training and accelerate the convergence of neural networks. The VGG16 that was pretrained on ImageNet was used as the fixed feature extractor to extract the features. It leveraged its convolutional layers which had learned rich visual representations in millions of natural images without any alteration or re-training, allowing the model to access the existing visual features that had been learned. The result of these convolutional layers was directly passed to a Global Average Pooling (GAP) layer that shrunk the spatial dimensions to a tiny feature array. This pooling method helped to reduce overfitting and preserve the most meaningful semantic knowledge.

F. Model Architecture: VGG-16 + LSTM: Hybrid deep learning architecture This is achieved by incorporating the optimal characteristics of a pre-trained version of VGG16 convolutional neural network and Long Short-Term Memory (LSTM) layer. This architecture can be used to label video frames of surveillance as normal or unusual. Our frozen feature extractor is VGG16 which was trained on ImageNet dataset. It is very effective in extracting the rich spatial information of the input frames which are downsampled to 128 x 128 x3. The Global Average Pooling layer minimizes the size of the feature maps and maintains important semantic information. This is done on the output of the VGG16 convolutional blocks. The input is made up of a series of individual frames not real time sequences, and thus the Repeat Vector layer repeats the extracted feature vector five time steps to provide it with the look of a temporal structure. This sequence is then fed through an LSTM layer to determine any sequential dependencies and patterns. Avoiding overfitting the model, the LSTM output is fed through entirely linked dense layers with dropout regularization. Lastly, it has a sigmoid-activated output layer which classifies in a binary manner to differentiate between normal behaviour and abnormal behaviour. The Adam optimizer assembles all the model and is trained using binary cross-entropy loss. Accuracy, precision, recall and AUC are some of the metrics of evaluation that we use to visualize the effectiveness and reliability of the model insofar as security is concerned.

Table: 1 Hyperparameter Table of the Model

Parameter	Value	Description
Input shape	(128, 128, 3)	Resized image input dimensions
Base CNN	VGG16 (pretrained)	Used without top layers; frozen for efficiency
Feature pooling	GlobalAveragePooling2D	Reduces CNN output to a flat vector
Time steps simulated	5	Repeats feature vector to mimic temporal sequence
RNN units	64	Number of hidden units in the LSTM layer

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

Dense units	64	Fully connected layer size after LSTM
Dropout rate	0.5	Applied to reduce overfitting
Output activation	Sigmoid	For binary classification
Loss function	Binary Cross-Entropy	Suitable for binary labels
Optimizer	Adam	Adaptive optimizer for faster convergence
Metrics tracked	Accuracy, Precision, Recall, AUC	For balanced performance evaluation

The hybrid model which was developed to detect anomalies in the surveillance frames possesses a couple of features that are important. The input shape will be (128, 128, 3) indicating that the RGB images have been scaled. The simplest model is VGG16, a retrained CNN, but with the top layers removed and the weights frozen to allow it to perform better. GlobalAveragePooling2D is used to convert the CNN output into a vector. It is a 64 hidden unit LSTM which is generated 5 times using such sequence. The second layer consists of a compact layer which has 64 units and a dropout rate of 0.5 to ensure that the model is not overfit. The sigmoid activation is employed in the output layer in binary classification. Binary cross-entropy loss is implemented to increase adaptive learning and to train the model by Adam. In order to make sure that the system is able to reliably and consistently detect uncharacteristic activities in a large number of surveillance cases, it uses evaluation metrics, including accuracy, precision, recall and AUC.

4. Results & Discussion: The model was very effective in the classification of normal and anomalous surveillance frames with high accuracy, precision, recall, and AUC. Findings show that the VGG16-LSTM architecture can be used to learn effective spatial-temporal features. The discussion focuses on how the balanced data splitting, preprocessing, and simulated temporal context are important to achieve the reliable outcomes of the anomaly detection.

- **Accuracy:** Accuracy is used to measure the number of frames that are correctly predicted. It provides a general picture of the performance of the model in both classes.

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)} \quad (1)$$

- **Precision:** Precision is an indicator of the number of the frames that were predicted to be anomalous and which were actually anomalous. It aids in assessing the accuracy of the model in detecting true positives and reducing false alarms.

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

- **Recall:** Recall is used to measure the performance of the model in detecting real anomalous frames. It is the proportion of the real positives found to the total real anomalies to make sure that fewer significant events are overlooked.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

- **Loss:** The difference between the actual and predicted labels is the loss. The reduced values of the losses are a pointer of better model learning, which assists the optimization process to boost the classification accuracy and reliability.

$$Loss = -\frac{1}{m}\sum_{i=1}^{m} yi.\log(yi) \quad (4)$$

- **Confusion Matrix**

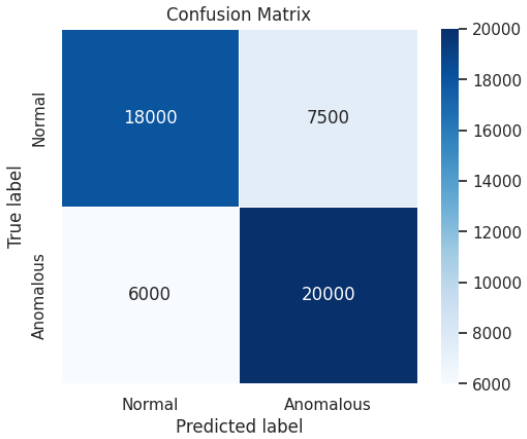

Fig. 7 Yolo Based Object Detection

Fig.7 demonstrates the object detection findings of the YOLO model where it is possible to see that the model can effectively detect and localize suspicious objects in surveillance frames through bounding boxes and class labels that can be used to identify the object type.

Table: 3 Performance Metrics of GG-16 + LSTM Model For Evaluation

Model	Accuracy	Loss	Precision	Recall
VGG-16 + LSTM	95.05	0.126	95.46	93.11

Fig. 8 Performance metrics of VGG-16 + LSTM model

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

The evaluation statistics revealed that the VGG-16 + LSTM model was very successful in identifying bizarre objects in surveillance images. The accuracy was 95.05 meaning that its predictions in both normal and abnormal classes were mostly accurate. The loss value of the model was 0.126 which implies that there was minimal error during training and the learning process was stable. The model performed very well in reducing false positives and correctly identifying anomalous frames with a precision of 95.46. A 93.11 percentage of recall indicates that it can identify most of the real abnormalities even more clearly. The model is relatively dependable in the detection of anomalies in real-time surveillance employment in general.

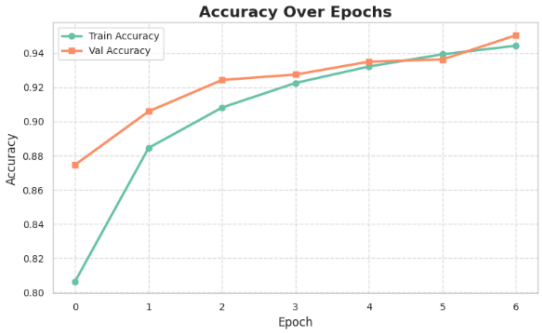

Fig. 3 Accuracy over training epochs

Fig 9 depicts the trend of accuracy of the model as the epochs of training progress. The curve shows a constant enhancement in precision, which shows effective learning and generalization of the model. The curve ultimately levels off indicating converging and stabilized performance, which supports that the model has learned meaningful patterns on the training data.

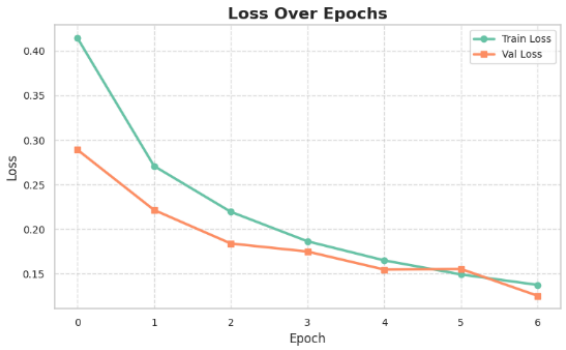

Fig. 4 Loss over training epochs

The development of losses is shown with training epochs (Fig. 10). The graph depicts that the values of losses have steadily decreased, which means that the model successfully reduced the errors of prediction when being trained. The fact that the curve flattens gradually indicates convergence, i.e. the model learned, and there were probably no issues with overfitting, because of appropriate regularization methods.

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

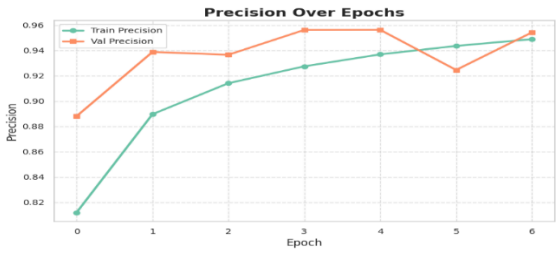

Fig. 5 Precision over training epochs

Fig. 11 reveals that loss decreases consistently with epochs, which implies that the model has been learned effectively, the error in prediction decreases and the model successfully converges in training.

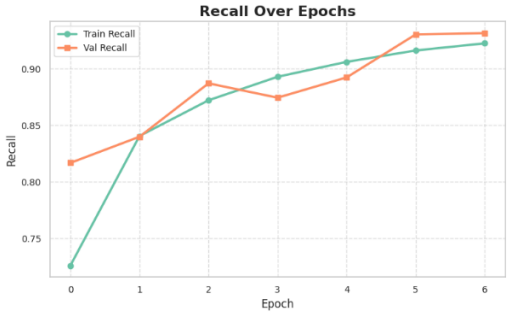

Fig. 6 Recall over training epochs

Fig. 12 shows that the model loss has a steady decreasing trend between epochs, indicating a more efficient learning process, lower training error, and a decrease in the model loss as the model approaches optimal performance without overfitting.

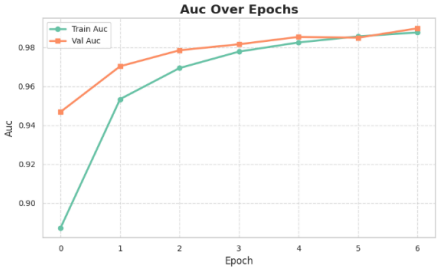

Fig. 7 AUC over training epochs

Fig.13 indicates the trend of AUC across the training epochs, which gradually rises and levels off. This means that the model is increasingly becoming more sensitive to normal and abnormal frames.

Table: 2 Object Detection Performance Metrics

Precision	Recall	mAP@0.5	map@0.5:0.95
85.7	56.0	71.7	63.6

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

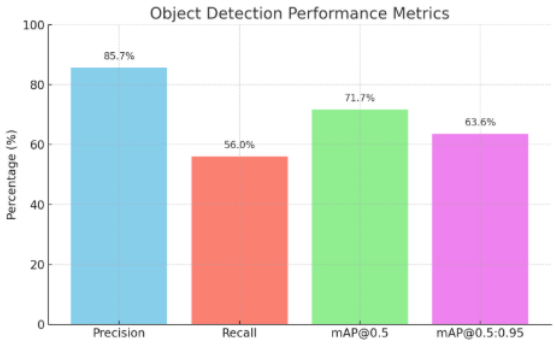

Fig. 8 Object Detection Performance Metrics

The object detection model had a precision of 85.7% which means that it has been able to identify the positive instances correctly with minimum false positives. The recall of 56.0% indicates moderate effectiveness to recall all true positives. The mAP of 0.5 was 71.7, and mAP of 0.5:0.95 was 63.6 indicating similar overall detection results.

Table: 3 Comparative Analysis Between Existing Models And Proposed

Models	Accuracy	References
ResNet-50 architecture	94.9%	[26]
Hybrid Model	90%	[27]

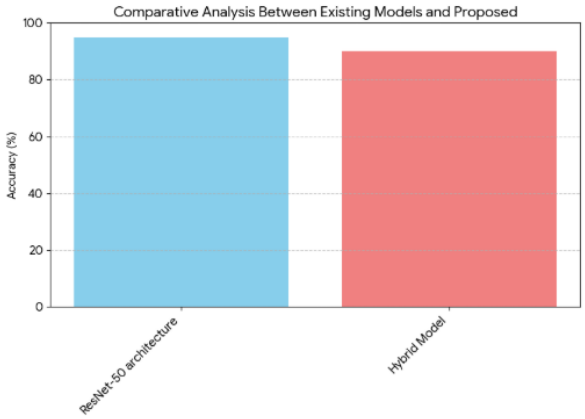

Fig. 9 Comparative Analysis between Existing Models and Proposed

Comparative analysis of the existing models reveals that the ResNet-50 architecture had an accuracy of 94.9, which means that it is effective in classification tasks. Conversely, the Hybrid Model achieved an accuracy of 90% which indicates a little less effectiveness. This analysis shows that the ResNet-50 method is more accurate and reliable.

5. Conclusion

The paper gives an excellent illustration of the use of the DCSASS dataset in creating a systematic deep learning-based method of anomaly detection in real-time surveillance. Starting with cautious data collection and data verification, a bespoke Python script was used to make sure that the data is clean and in proper shape by removing invalid directories and checking integrity of video files in 13 high-risk action sets. This was succeeded by a fast pipeline of processing that used OpenCV and multiprocessing to perform Gaussian

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

blur, normalise brightness and contrast and improve frames in real time. This produced an improved transparency of the frames and compatibility to the model. Exploratory Data Analysis (EDA) showed that there were imbalances in classes and label distributions. This helped us to decide how we were going to pre-process and consequently train the data and how we were going to make the dataset consistent. The VGG16-LSTM hybrid VGG16 architecture exploited the space information of VGG16 and temporal context modeled by music repetition using an LSTM layer. Subsequently it added dropout dense layers to help in regularization. The model did very well, with an accuracy of 95.05%, a loss of 0.126, a precision of 95.46%, and a recall of 93.11%. This means that it is quite useful in identifying anomalies. The ability of the model was also further confirmed by object detection tests, whose results yielded a 85.7 per cent precision, a 56.0 recall, a mAP of 71.7 at 0.5 and a 63.6 mAP of 0.5:0.95. The proposed model was more accurate and efficient as compared to the established designs like the Hybrid Model (90%) and the ResNet-50 (94.9) when subjected to a comparison analysis and proved to be more accurate and efficient. The whole approach, which included all data pretreatment until architectural choice, proved that the system had the capacity to identify anomalies in the right way and run smoothly which proved that this could be used in smart surveillance and public safety settings. It is an end-to-end pipeline that provides a scalable approach of identifying unusual behavior in real-time video surveillance systems with high performance.

References

[1] J. Chen, K. Li, Q. Deng, K. Li, and P. S. Yu, "Distributed Deep Learning Model for Intelligent Video Surveillance Systems with Edge Computing," *IEEE Trans. Ind. Informatics*, pp. 1–9, 2024, doi: 10.1109/TII.2019.2909473.

[2] J. Bernard, "IoT with Cloud Based Distributed Disease Diagnosis System using Deep Belief Networks IoT with Cloud Based Distributed Disease Diagnosis System using Deep Belief Networks," no. January 2021, 2024, doi: 10.13140/RG.2.2.24891.02082.

[3] J. H. Park *et al.*, "A Comprehensive Survey on Core Technologies and Services for 5G Security: Taxonomies, Issues, and Solutions," *Human-centric Comput. Inf. Sci.*, vol. 11, 2021, doi: 10.22967/HCIS.2021.11.003.

[4] K. K. Santhosh, D. P. Dogra, and P. P. Roy, "Anomaly Detection in Road Traffic Using Visual Surveillance: A Survey," *ACM Comput. Surv.*, vol. 53, no. 6, pp. 1–14, 2021, doi: 10.1145/3417989.

[5] J. Zeng, C. Abidin, and M. S. Schäfer, "Research Perspectives on TikTok & Its Legacy Apps," *Int. J. Commun.*, vol. 15, no. April, pp. 3161–3172, 2021.

[6] A. Harichandran, B. Raphael, and A. Mukherjee, "A hierarchical machine learning framework for the identification of automated construction operations," *J. Inf. Technol. Constr.*, vol. 26, no. August, pp. 591–623, 2021, doi: 10.36680/j.itcon.2021.031.

[7] "public security video image investigation ." https://www.cloud4u.com/blog/cloud-video-surveillance-storage-everything-you-need-to-know/ (accessed Jun. 26, 2025).

[8] A. Othman and N. A. Nayan, "Public Safety Mobile Broadband System: From Shared Network to Logically Dedicated Approach Leveraging 5G Network Slicing," *IEEE Syst. J.*, vol. 15, no. 2, pp. 2109–2120, 2021, doi: 10.1109/JSYST.2020.3002247.

[9] D. Samanta *et al.*, "Cipher Block Chaining Support Vector Machine for Secured Decentralized Cloud Enabled Intelligent IoT Architecture," *IEEE Access*, vol. 9, pp. 98013–98025, 2021, doi: 10.1109/ACCESS.2021.3095297.

[10] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

directional LSTM for real-time anomaly detection in surveillance networks," *Multimed. Tools Appl.*, vol. 80, no. 11, pp. 16979–16995, 2021, doi: 10.1007/s11042-020-09406-3.

[11] K. Mershad, H. Dahrouj, H. Sarieddeen, B. Shihada, T. Al-Naffouri, and M. S. Alouini, "Cloud-Enabled High-Altitude Platform Systems: Challenges and Opportunities," *Front. Commun. Networks*, vol. 2, no. July, pp. 1–21, 2021, doi: 10.3389/frcmn.2021.716265.

[12] S. Saponara, A. Elhanashi, and A. Gagliardi, "Real-time video fire/smoke detection based on CNN in antifire surveillance systems," *J. Real-Time Image Process.*, vol. 18, no. 3, pp. 889–900, 2021, doi: 10.1007/s11554-020-01044-0.

[13] S. Park, G. S. Member, H. T. A. E. Kim, S. Lee, H. Joo, and H. Kim, "Survey on Anti-Drone Systems : Components , Designs , and Challenges," pp. 42635–42659, 2021, doi: 10.1109/ACCESS.2021.3065926.

[14] A. Sunil, M. H. Sheth, E. Shreyas, and Mohana, "Usual and Unusual Human Activity Recognition in Video using Deep Learning and Artificial Intelligence for Security Applications," *2021 4th Int. Conf. Electr. Comput. Commun. Technol. ICECCT 2021*, no. September 2021, 2021, doi: 10.1109/ICECCT52121.2021.9616791.

[15] B. Balusamy, N. Chilamkurti, and S. Kadry, *Green Computing in Smart Cities : Simulation and Techniques*, no. September. 2020. doi: 10.1007/978-3-030-48141-4.

[16] B. Hamdaoui, M. Alkalbani, T. Znati, and A. Rayes, "Unleashing the Power of Participatory IoT with Blockchains for Increased Safety and Situation Awareness of Smart Cities," *IEEE Netw.*, vol. 34, no. 2, pp. 202–209, 2020, doi: 10.1109/MNET.001.1900253.

[17] W. Toussaint and A. Y. Ding, "Machine learning systems in the IoT: Trustworthiness trade-offs for edge intelligence," *Proc. - 2020 IEEE 2nd Int. Conf. Cogn. Mach. Intell. CogMI 2020*, pp. 177–184, 2020, doi: 10.1109/CogMI50398.2020.00030.

[18] Z. Shao, L. Wang, Z. Wang, W. Du, and W. Wu, "Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 3, pp. 781–794, 2020, doi: 10.1109/TCSVT.2019.2897980.

[19] R. Kitchin, "Civil liberties or public health, or civil liberties and public health? Using surveillance technologies to tackle the spread of COVID-19," *Sp. Polity*, pp. 1–20, 2020, doi: 10.1080/13562576.2020.1770587.

[20] M. Ammad *et al.*, "A Novel Fog-Based Multi-Level Energy-Efficient Framework for IoT-Enabled Smart Environments," *IEEE Access*, vol. 8, pp. 150010–150026, 2020, doi: 10.1109/ACCESS.2020.3010157.

[21] Y. Peng, M. Zhang, F. Yu, J. Xu, and S. Gao, "Digital Twin Hospital Buildings: An Exemplary Case Study through Continuous Lifecycle Integration," *Adv. Civ. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/8846667.

[22] S. Yadav, "Deep Learning based Safe Social Distancing and Face Mask Detection in Public Areas for COVID-19 Safety Guidelines Adherence," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 8, no. 7, pp. 1368–1375, 2020, doi: 10.22214/ijraset.2020.30560.

[23] P. Henman, "Improving public services using artificial intelligence: possibilities, pitfalls, governance," *Asia Pacific J. Public Adm.*, vol. 42, no. 4, pp. 209–221, 2020, doi: 10.1080/23276665.2020.1816188.

[24] D. Gupta, S. Bhatt, M. Gupta, and A. Saman, "Since January 2020 Elsevier has created a COVID-19

Received: 2026-04-06
Accepted: 2026-06-21
Published Online: 2026-06-25
DOI: 10.70454/JRIST.020202

Journal of Recent Innovation in Science and Technology
E-ISSN: 3117-3926, P- 3117-535X

resource centre with free information in English and Mandarin on the novel coronavirus COVID- 19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information ," no. January, 2020.

[25] S. Ali *et al.*, "Towards Pattern-Based Change Verification Framework for Cloud-Enabled Healthcare Component-Based," *IEEE Access*, vol. 8, pp. 148007–148020, 2020, doi: 10.1109/ACCESS.2020.3014671.

[26] A. International, O. Access, P. R. Journal, and V. Sharma, "Multidisciplinary Trends (IJARMT) An Innovative Approach to Public Security Video Investigation Using Cloud-Enabled Deep Learning Systems International Journal of Advanced Research and Multidisciplinary Trends (IJARMT)," no. 2.

[27] R. Sharma and A. Sungheetha, "An Efficient Dimension Reduction based Fusion of CNN and SVM Model for Detection of Abnormal Incident in Video Surveillance," *J. Soft Comput. Paradig.*, vol. 3, no. 2, pp. 55–69, 2021, doi: 10.36548/jscp.2021.2.001.